AMENDMENTS TO THE SPECIFICATION

Please insert the following paragraph after paragraph [0001]:

SEQUENCE LISTING COMPACT DISK APPENDIX

[0001.1] Two copies of compact discs labeled Copy 1 and Copy 2 of the Sequence Listing are attached to this application. Each compact disc contains a single file, SEQLIST506612000100.txt (created on May 3, 2006, 1.55 MB), the contents of which are hereby incorporated by reference.

Please replace paragraph [0098] on page 19 with the following amended paragraph:

[0098] In the case of mammalian genomes, 2C = ~3.6 X 109 , and an oligonucleotide of 14-15 nucleotides is expected to be represented only once in the genome. However, the distribution of nucleotides in the coding sequence of mammalian genomes is nonrandom (Lathe, R. J. Mol. Biol. 183:1 (1985) and longer oligonucleotides may be preferred in order to in increase the specificity of hybridization. In practical terms, this works out to probes that are 19-40 nucleotides long (Sambrook J et al., infra). The second method for estimating the length of a specific probe is to use a probe long enough to hybridize under the chosen conditions and use a computer to search for that sequence or close matches to the sequence in the human genome and choose a unique match. Probe sequences are chosen based on the desired hybridization properties as described in Chapter 11 of Sambrook et al, infra. The PRIMER3 program is useful for designing these probes (S. Rozen and H. Skaletsky 1996,1997; Primer3 code available at ~~http://www-genome.wi.mit.edu/genome_software/other/primer3.html~~www-genome.wi.mit.edu/genome_software/other/primer3.html). The sequences of these probes are then compared pair wise against a database of the human genome sequences using a program such as BLAST or MEGABLAST (Madden, T.L et al.(1996) Meth. Enzymol. 266:131-141). Since most of the human genome is now contained in the database, the number of matches will be determined. Probe sequences are chosen that are unique to the desired target sequence.

Please replace paragraph [0101] on page 21 with the following amended paragraph:

[0101] Similarly, commercial sources for nucleic acid and protein microarrays are available, and include, e.g., Agilent Technologies, Palo Alto, CA (~~http://www.agilent.com/~~www.agilent.com) Affymetrix, Santa Clara,CA (~~http://www.affymetrix.com/~~www.affymetrix.com); and Incyte, Palo Alto, CA (~~http://www.incyte.com/~~www.incyte.com) and others.

Please replace paragraph [0106] on page 23 with the following amended paragraph:

[0106] Firstly, publication and sequence databases can be "mined" using a variety of search strategies, including, e.g., a variety of genomics and proteomics approaches. For example, currently available scientific and medical publication databases such as Medline, Current Contents, OMIM (online Mendelian inheritance in man) various Biological and Chemical Abstracts, Journal indexes, and the like can be searched using term or key-word searches, or by author, title, or other relevant search parameters. Many such databases are publicly available, and one of skill is well versed in strategies and procedures for identifying publications and their contents, e.g., genes, other nucleotide sequences, descriptions, indications, expression pattern, etc. Numerous databases are available through the internet for free or by subscription, *see*, e.g., ~~http://www.ncbi.nlm.nih.gov/PubMed/~~www.ncbi.nlm.nih.gov/PubMed; ~~http://www3.infotrieve.com/~~www3.infotrieve.com; ~~http://www.isinet.com/~~www.isinet.com; ~~http://www.sciencemag.org/~~www.sciencemag.org. Additional or alternative publication or citation databases are also available that provide identical or similar types of information, any of which are favorable employed in the context of the invention. These databases can be searched for publications describing differential gene expression in leukocytes between patient with and without diseases or conditions listed in Table 1. We identified the nucleotide sequences listed in Table 2 and some of the sequences listed in Table 8 (Example 20), using data mining methods.

Please replace paragraph [0107] on page 23 with the following amended paragraph:

[0107] Alternatively, a variety of publicly available and proprietary sequence databases (including GenBank, dbEST, UniGene, and TIGR and SAGE databases) including sequences corresponding to expressed nucleotide sequences, such as expressed sequence tags (ESTs) are available. For example, Genbank™ (~~http://www.ncbi.nlm.nih.gov/Genbank/~~

www.ncbi.nlm.nih.gov/Genbank) among others can be readily accessed and searched via the internet. These and other sequence and clone database resources are currently available; however, any number of additional or alternative databases comprising nucleotide sequence sequences, EST sequences, clone repositories, PCR primer sequences, and the like corresponding to individual nucleotide sequence sequences are also suitable for the purposes of the invention. Sequences from nucleotide sequences can be identified that are only found in libraries derived from leukocytes or sub-populations of leukocytes, for example see Table 2.

Please replace paragraph [0145] on page 36 with the following amended paragraph:

[0145] Alternatively, expression at the level of protein products of gene expression is performed. For example, protein expression, in a sample of leukocytes, can be evaluated by one or more method selected from among: western analysis, two-dimensional gel analysis, chromatographic separation, mass spectrometric detection, protein-fusion reporter constructs, colorimetric assays, binding to a protein array and characterization of polysomal mRNA. One particularly favorable approach involves binding of labeled protein expression products to an array of antibodies specific for members of the candidate library. Methods for producing and evaluating antibodies are widespread in the art, *see*, e.g., Coligan, *supra*; and Harlow and Lane (1989) Antibodies: A Laboratory Manual, Cold Spring Harbor Press, NY ("Harlow and Lane"). Additional details regarding a variety of immunological and immunoassay procedures adaptable to the present invention by selection of antibody reagents specific for the products of candidate nucleotide sequences can be found in, e.g., Stites and Terr (eds.)(1991) Basic and Clinical Immunology, 7th ed., and Paul, *supra*. Another approach uses systems for performing desorption spectrometry. Commercially available systems, e.g., from Ciphergen Biosystems, Inc. (Fremont, CA) are particularly well suited to quantitative analysis of protein expression. Indeed, Protein Chip® arrays (*see*, e.g., ~~http://www.ciphergen.com/~~www.ciphergen.com) used in desorption spectrometry approaches provide arrays for detection of protein expression. Alternatively, affinity reagents, e.g., antibodies, small molecules, etc.) are developed that recognize epitopes of the protein product. Affinity assays are used in protein array assays, e.g. to detect the presence or absence of particular proteins. Alternatively, affinity reagents are used to detect expression using the methods described above. In the case of a protein that is expressed on the cell surface of leukocytes, labeled affinity

reagents are bound to populations of leukocytes, and leukocytes expressing the protein are identified and counted using fluorescent activated cell sorting (FACS).

Please replace paragraph [0327] on page 95 with the following amended paragraph:

[0327] Software for performing the statistical methods required for the invention, e.g., to determine correlations between expression profiles and subsets of members of the diagnostic nucleotide libraries, such as programmed embodiments of the statistical methods described above, are also included in the computer systems of the invention. Alternatively, programming elements for performing such methods as principle component analysis (PCA) or least squares analysis can also be included in the digital system to identify relationships between data. Exemplary software for such methods is provided by Partek, Inc., St. Peter, Mo; ~~http://www.partek.com~~www.partek.com.

Please replace paragraph [0359] on page 107 with the following amended paragraph:

[0359] Next, two publicly available databases of DNA sequences, Unigene (~~http://www.ncbi.nlm.nih.gov/UniGene/~~www.ncbi.nlm.nih.gov/UniGene) and BodyMap (~~http://bodymap.ims.u-tokyo.ac.jp/~~bodymap.ims.u-tokyo.ac.jp), were searched for sequenced DNA clones that showed specificity to leukocyte lineages, or subsets of leukocytes, or resting or activated leukocytes.

Please replace paragraph [0360] on page 107 with the following amended paragraph:

[0360] The human Unigene database (build 133) was used to identify leukocyte candidate nucleotide sequences that were likely to be highly or exclusively expressed in leukocytes. We used the Library Differential Display utility of Unigene (~~http://www.ncbi.nlm.nih.gov/UniGene/info/ddd.html~~www.ncbi.nlm.nih.gov/UniGene/info/ddd.html), which uses statistical methods (The Fisher Exact Test) to identify nucleotide sequences that have

relative specificity for a chosen library or group of libraries relative to each other. We compared the following human libraries from Unigene release 133:

546	NCI_CGAP_HSC1 (399)
848	Human_mRNA_from_cd34+_stem_cells (122)
105	CD34+DIRECTIONAL (150)
3587	KRIBB_Human_CD4_intrathymic_T-cell_cDNA_library (134)
3586	KRIBB_Human_DP_intrathymic_T-cell_cDNA_library (179)
3585	KRIBB_Human_TN_intrathymic_T-cell_cDNA_library (127)
3586	323 Activated_T-cells_I (740)
376	Activated_T-cells_XX (1727)
327	Monocytes,_stimulated_II (110)
824	Proliferating_Erythroid_Cells_(LCB:ad_library) (665)
825	429 Macrophage_II (105)
387	Macrophage_I (137)
669	NCI_CGAP_CLL1 (11626)
129	Human_White_blood_cells (922)
1400	NIH_MGC_2 (422)
55	Human_promyelocyte (1220)
1010	NCI_CGAP_CML1 (2541)
2217	NCI_CGAP_Sub7 (218)
1395	NCI_CGAP_Sub6 (2764)
4874	NIH_MGC_48 (2524)

Please replace paragraph [0363] on page 107 with the following amended paragraph:

[0363] DNA clones corresponding to each UniGene cluster number are obtained in a variety of ways. First, a cDNA clone with identical sequence to part of, or all of the identified UniGene cluster is bought from a commercial vendor or obtained from the IMAGE consortium (~~http://image.llnl.gov/~~image.llnl.gov, the Integrated Molecular Analysis of Genomes and their

Expression). Alternatively, PCR primers are designed to amplify and clone any portion of the nucleotide sequence from cDNA or genomic DNA using well-known techniques. Alternatively, the sequences of the identified UniGene clusters are used to design and synthesize oligonucleotide probes for use in microarray based expression profiling.

Please replace paragraph [0368] on page 110 with the following amended paragraph:

[0368] Messenger RNA contains repetitive elements that are found in genomic DNA. These repetitive elements lead to false positive results in similarity searches of query mRNA sequences versus known mRNA and EST databases. Additionally, regions of low information content (long runs of the same nucleotide, for example) also result in false positive results. These regions were masked using the program RepeatMasker2 ~~found at http://repeatmasker.genome.washington.edu (Smit, AFA & Green, P "RepeatMasker" at http://ftp.genome.washington.edu/RM/RepeatMasker.html~~repeatmasker.genome.washington.edu (Smit, AFA & Green, P "RepeatMasker" at genome.washington.edu/RM/RepeatMasker.html). The trimmed and masked files were then subjected to further sequence analysis.

Please replace paragraph [0369] on page 111 with the following amended paragraph:

[0369] cDNA sequences were further characterized using BLAST analysis. The BLASTN program was used to compare the sequence of the fragment to the UniGene, dbEST, and nr databases at NCBI (GenBank release 123.0; see Table 5). In the BLAST algorithm, the expect value for an alignment is used as the measure of its significance. First, the cDNA sequences were compared to sequences in Unigene (~~http://www.ncbi.nlm.nih.gov/UniGene~~ www.ncbi.nlm.nih.gov/UniGene). If no alignments were found with an expect value less than 10^{-25}, the sequence was compared to the sequences in the dbEST database using BLASTN. If no alignments were found with an expect value less than 10^{-25}, the sequence was compared to sequences in the nr database.

Please replace paragraph [0370] on page 111 with the following amended paragraph:

[0370] The BLAST analysis produced the following categories of results: a) a significant match to a known or predicted human gene, b) a significant match to a nonhuman DNA sequence, such as vector DNA or *E. coli* DNA, c) a significant match to an unidentified GenBank entry (a sequence not previously identified or predicted to be an expressed sequence or a gene), such as a cDNA clone, mRNA, or cosmid , or d) no significant alignments. If a match to a known or predicted human gene was found, analysis of the known or predicted protein product was performed as described below. If a match to an unidentified GenBank entry was found, or if no significant alignments were found, the sequence was searched against all known sequences in the human genome database (~~http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs~~ www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs, see Table 5).

Please replace paragraph [0372] on page 112 with the following amended paragraph:

[0372] In some cases, the process of analyzing many unknown sequences with BLASTN was automated by using the BLAST network-client program blastcl3, which was downloaded via ftp from ~~ftp://ncbi.nlm.nih.gov/blast/network/netblast~~ ncbi.nlm.nih.gov/blast/network/netblast.

Please replace paragraph [0377] on page 113 with the following amended paragraph:

[0377] This sequence was used as input for a series of BLASTN searches. First, it was used to search the UniGene database, build 132 (~~http://www.ncbi.nlm.nih.gov/BLAST/~~ www.ncbi.nlm.nih.gov/BLAST). No alignments were found with an expect value less than the threshold value of 10^{-25}. A BLASTN search of the database dbEST, release 041001, was then performed on the sequence and 21 alignments were found (~~http://www.ncbi.nlm.nih.gov/BLAST/~~ www.ncbi.nlm.nih.gov/BLAST). Ten of these had expect values less than 10^{-25}, but all were matches to unidentified cDNA clones. Next, the sequence was used to run a BLASTN search of the

nr database, release 123.0. No significant alignment to any sequence in nr was found. Finally, a BLASTN search of the human genome was performed on the sequence (~~http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs~~ www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs).

Please replace paragraph [0378] on page 114 with the following amended paragraph:

[0378] A single alignment to the genome was found on contig NT_004698.3 (e=0.0). The region of alignment on the contig was from base 1,821,298 to base 1,822,054, and this region was found to be mapped to chromosome 1, from base 105,552,694 to base 105,553,450. The sequence containing the aligned region, plus 100 kilobases on each side of the aligned region, was downloaded. Specifically, the sequence of chromosome 1 from base105,452,694 to 105,653,450 was downloaded (~~http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/seq_reg.cgi?chr=1&from=105452694&to=105653450~~www.ncbi.nlm.nih.gov/cgi-bin/Entrez/seq_reg.cgi?chr=1&from=105452694&to=105653450).

Please replace paragraph [0379] on page 114 with the following amended paragraph:

[0379] This 200,757 bp segment of the chromosome was used to predict exons and their peptide products as follows. The sequence was used as input for the Genscan algorithm (~~http://genes.mit.edu/GENSCAN.html~~genes.mit.edu/GENSCAN.html), using the following Genscan settings:

Please replace paragraph [0385] on page 115 with the following amended paragraph:

[0385] At least 100 significant alignments were found in the nr database, as well. A similarity to hypothetical protein FLJ22457 (UniGene cluster Hs.238707)was found (e=0.0). The cDNA of this predicted protein has been isolated from B lymphocytes (~~http://www.ncbi.nlm.nih.gov/entrez/viewer.cgi?save=0&cmd=&cfm=on&f=1&view=gp&txt=0&v~~

~~al=13637988~~www.ncbi.nlm.nih.gov/entrez/viewer.cgi?save=0&cmd=&cfm=on&f=1&view=gp&txt=0&val=13637988).

Please replace paragraph [0388] on page 115 with the following amended paragraph:

[0388] Multiple analyses were performed using this prediction. First, a pairwise comparison of the sequence above and the sequence of FLJ22457, the hypothetical protein mentioned above, using BLASTP version 2.1.2 (~~http://ncbi.nlm.nih.gov/BLAST/~~ ncbi.nlm.nih.gov/BLAST), resulted in a match with an expect value of 0.0. The peptide sequence predicted from clone 596H6 was longer and 19% of the region of alignment between the two resulted from gaps in hypothetical protein FLJ22457. The cause of the discrepancy might be alternative mRNA splicing, alternative post-translational processing, or differences in the peptide-predicting algorithms used to create the two sequences, but the homology between the two is significant.

Please replace paragraph [0391] on page 116 with the following amended paragraph:

[0391] To discover similarities to protein families, comparisons of the domains (described above) were carried out using the Pfam and Blocks databases. A search of the Pfam database identified two regions of the peptide domains as belonging the DENN protein family ($e=2.1 \times 10^{-33}$). The human DENN protein possesses an RGD cellular adhesion motif and a leucine-zipper-like motif associated with protein dimerization, and shows partial homology to the receptor binding domain of tumor necrosis factor alpha. DENN is virtually identical to MADD, a human MAP kinase-activating death domain protein that interacts with type I tumor necrosis factor receptor (~~http://srs.ebi.ac.uk/srs6bin/cgi-bin/wgetz?-id+fS5n1GQsHf+-e+[INTERPRO:'IPR001194']~~srs.ebi.ac.uk/srs6bin/cgi-bin/wgetz?-id+fS5n1GQsHf+-e+[INTERPRO:'IPR001194']). The search of the Blocks database also revealed similarities between regions of the peptide sequence and known protein groups, but none with a satisfactory degree of confidence. In the Blocks scoring system, scores over 1,100 are likely to be relevant. The highest score of any match to the predicted peptide was 1,058.

Please replace paragraph [0395] on page 117 with the following amended paragraph:

[0395] Membrane-spanning regions were predicted by graphing hydrophobicity vs. amino acid number. Thirteen regions were found to be somewhat hydrophobic. The algorithm TMpred predicted a model with 6 strong transmembrane helices (~~http://www.ch.embnet.org/software/TMPRED_form.html~~www.ch.embnet.org/software/TMPRED_form.html).

Please replace paragraph [0401] on page 119 with the following amended paragraph:

[0401] The sequence of the CAP2 contig was used in a BLAST search of the human genome. 934 out of 1,010 residues aligned to a region of chromosome 21. A gap of 61 residues divided the aligned region into two smaller fragments. The sequence of this region, plus 100 kilobases on each side of it, was downloaded and analyzed using the Genscan site at MIT (~~http://genes.mit.edu/GENSCAN.html~~genes.mit.edu/GENSCAN.html), with the following settings:

Please replace paragraph [0406] on page 119 with the following amended paragraph:

[0406] The peptide sequence predicted by Genscan was also saved. Multiple types of analyses were performed on it using the resources mentioned in Table 3. BLASTP and TBLASTN were used to search the TrEMBL protein database (~~http://www.expasy.ch/sprot~~ www.expasy.ch/sprot) and the GenBank nr database (~~http://www.ncbi.nlm.hih.gov/BLAST~~ www.ncbi.nlm.hih.gov/BLAST), which includes data from the SwissProt, PIR, PRF, and PDB databases. No significant matches were found in any of these, so no gene identity or tertiary structure was discovered.

Please replace paragraph [0415] on page 121 with the following amended paragraph:

[0415] Spotted cDNA microarrays were then made from these PCR products by ArrayIt using their protocols (~~http://arrayit.com/Custom_Microarrays/Flex-Chips/flex-chips.html~~ arrayit.com/Custom_Microarrays/Flex-Chips/flex-chips.html). Each fragment was spotted 3 times onto each array.

Please replace paragraph [0417] on page 122 with the following amended paragraph:

[0417] Oligonucleotide probes are also prepared using the DNA sequence information for the candidate genes identified by differential hybridization screening (listed in Table 3 and the sequence listing) and/or the sequence information for the genes identified by database mining (listed in Table 2) is used to design complimentary oligonucleotide probes. Oligo probes are designed on a contract basis by various companies (for example, Compugen, Mergen, Affymetrix, Telechem), or designed from the candidate sequences using a variety of parameters and algorithms as indicated at ~~http://www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi~~ www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi. Briefly, the length of the oligonucleotide to be synthesized is determined, preferably greater than 18 nucleotides, generally 18-24 nucleotides, 24-70 nucleotides and, in some circumstances, more than 70 nucleotides. The sequence analysis algorithms and tools described above are applied to the sequences to mask repetitive elements, vector sequences and low complexity sequences. Oligonucleotides are selected that are specific to the candidate nucleotide sequence (based on a Blast n search of the oligonucleotide sequence in question against gene sequences databases, such as the Human Genome Sequence, UniGene, dbEST or the non-redundant database at NCBI), and have <50% G content and 25–70% G+C content. Desired oligonucleotides are synthesized using well-known methods and apparatus, or ordered from a company (for example Sigma). Oligonucleotides are spotted onto microarrays. Alternatively, oligonucleotides are synthesized directly on the array surface, using a variety of techniques (Hughes et al. 2001, Yershov et al. 1996, Lockhart et al 1996).

Please replace paragraph [0558] on page 153 with the following amended paragraph:

[0558] Probes were designed from database sequences that had the highest similarity to each of the sequenced clones in Tables 3A, 3B, and 3C. Based on BLASTn searches the most similar database sequence was identified by locus number and the locus number was submitted to GenBank using batch Entrez (~~http://www.ncbi.nlm.nih.gov/entrez/batchentrez.cgi?db=Nucleotide~~ www.ncbi.nlm.nih.gov/entrez/batchentrez.cgi?db=Nucleotide) to obtain the sequence for that locus. The GenBank entry sequence was used because in most cases it was more complete or was derived from multi-pass sequencing and thus would likely have fewer errors than the single pass cDNA library sequences. When only UniGene cluster IDs were available for genes of interest, the respective sequences were extracted from the UniGene_unique database, build 137, downloaded via ftp from NCBI (~~ftp://ncbi.nlm.nih.gov/repository/UniGene/~~ncbi.nlm.nih.gov/repository/UniGene). This database contains one representative sequence for each cluster in UniGene.

Summary of BioCardia library clones used in probe design.

Table	**Sense Strand**	**Antisense Strand**	**Strand Undetermined**
Table 3A	3621	763	124
Table 3B	142	130	238
Table 3C	19	6	23
Totals	**3782**	**899**	**385**

Please replace paragraph [0561] on page 154 with the following amended paragraph:

[0561] Database mining was performed as described in Example 2. In addition, the Library Browser at the NCBI UniGene web site (~~http://www.ncbi.nlm.nih.gov/UniGene/lbrowse.cgi?ORG=Hs&DISPLAY=ALL~~www.ncbi.nlm.nih.gov/UniGene/lbrowse.cgi?ORG=Hs&DISPLAY=ALL) was used to identify genes that are specifically expressed in leukocyte cell populations. All expression libraries available at the time were examined and those derived from leukocytes were viewed individually. Each library viewed through the Library Browser at the UniGene web site contains a section titled "Shown below are UniGene clusters of special interest only" that lists genes that are either highly represented or found

only in that library. Only the genes in this section were downloaded from each library. Alternatively, every sequence in each library is downloaded and then redundancy between libraries is reduced by discarding all UniGene cluster IDs that are represented more than once. A total of 439 libraries were downloaded, containing 35,819 genes, although many were found in more than one library. The most important libraries from the remaining set were separated and 3,914 genes remained. After eliminating all redundancy between these libraries and comparing the remaining genes to those listed in Tables 3A, 3B and 3C, the set was reduced to 2,573 genes in 35 libraries (listed below). From these, all genes in first 30 libraries were used to design probes. A random subset of genes was used from Library Lib.376, "Activated_T-cells_XX". From the last four libraries, a random subset of sequences listed as "ESTs, found only in this library" was used.

Library ID	**Library Name**	Category	**No. of sequences before reduction**	**No. of sequences used on array***
Lib.2228	Human_leukocyte_MATCHMAKER_cDNA_Library	other/unclassified	4	3
Lib.238	RA-MO-III (activated monocytes from RA patient)	Blood	2	1
Lib.242	Human_peripheral_blood_(Whole)_(Steve_Elledge)	Blood	4	2
Lib.2439	Subtracted_cDNA_libraries_from_human_Jurkat_cells	other/unclassified	4	1
Lib.323	Activated_T-cells_I	other/unclassified	19	3
Lib.327	Monocytes,_stimulated_II	Blood	92	35
Lib.387	Macrophage_I	other/unclassified	84	24
Lib.409	Activated_T-cells_IV	other/unclassified	37	10
Lib.410	Activated_T-cells_VIII	other/unclassified	27	10
Lib.411	Activated_T-cells_V	other/unclassified	41	9
Lib.412	Activated_T-cells_XII	other/unclassified	29	12
Lib.413	Activated_T-cells_XI	other/unclassified	13	6
Lib.414	Activated_T-cells_II	other/unclassified	69	30
Lib.429	Macrophage_II	other/unclassified	56	24
Lib.4480	Homo_sapiens_rheumatoid_arthritis_fibroblast-like_synovial	other/unclassified	7	6
Lib.476	Macrophage,_subtracted_(total_cDNA)	other/unclassified	11	1
Lib.490	Activated_T-cells_III	other/unclassified	9	5
Lib.491	Activated_T-cells_VII	other/unclassified	27	8

Lib.492	Activated_T-cells_IX	other/unclassified	16	5
Lib.493	Activated_T-cells_VI	other/unclassified	31	15
Lib.494	Activated_T-cells_X	other/unclassified	18	5
Lib.498	RA-MO-I (activated peripheral blood monocytes from RA patient)	Blood	2	1
Lib.5009	Homo_Sapiens_cDNA_Library_from_Peripheral_White_Blood_Cell	other/unclassified	3	3
Lib.6338	human_activated_B_lymphocyte	Tonsils	9	8
Lib.6342	Human_lymphocytes	other/unclassified	2	2
Lib.646	Human_leukocyte_(M.L.Markelov)	other/unclassified	1	1
Lib.689	Subtracted_cDNA_library_of_activated_B_lymphocyte	Tonsil	1	1
Lib.773	PMA-induced_HL60_cell_subtraction_library (leukemia)	other/unclassified	6	3
Lib.1367	cDNA_Library_from_rIL-2_activated_lymphocytes	other/unclassified	3	2
Lib.5018	Homo_sapiens_CD4+_T-cell_clone_HA1.7	other/unclassified	6	3
Lib.376	Activated_T-cells_XX	other/unclassified	999	119
Lib.669	NCI_CGAP_CLL1 (Lymphocyte)	Blood	353	81†
Lib.1395	NCI_CGAP_Sub6 (germinal center b-cells)	B cells germinal	389	100†
Lib.2217	NCI_CGAP_Sub7 (germinal center b-cells)	B cells germinal	605	200†
Lib.289	NCI_CGAP_GCB1 (germinal center b-cells)	Tonsil	935	200†
Total			**3,914**	**939**

* Redundancy of UniGene numbers between the libraries was eliminated.

† A subset of genes flagged as "Found only in this library" were taken.

Please replace paragraph [0582] on page 166 with the following amended paragraph:

[0582] The FASTA file, including the sequence of NM_002310, was masked using the RepeatMasker web interface (Smit, AFA & Green, P RepeatMasker at ~~http://ftp.genome.washington.edu/RM/RepeatMasker.html,~~ftp.genome.washington.edu/RM/RepeatMasker.html, Smit and Green). Specifically, during masking, the following types of sequences were replaced with "N's": SINE/MIR & LINE/L2, LINE/L1 , LTR/MaLR, LTR/Retroviral , Alu,

and other low informational content sequences such as simple repeats. Below is the sequence following masking:

On page 343, replace Table 5 with the following amended Table 5:

Table 5: Nucleotide sequence databases used for analysis

Database	**Version**	**<u>Description</u>**	**Location of file**	**Threshold of Significance Used**
nr	Release 123.0	GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, or HTGS sequences). No longer "non-redundant".	~~ftp:/~~/ncbi.nlm.nih.gov/ blast/nt.Z	Expect value (e) < 10^{-25}
dbEST	04/10/01	Non-redundant Database of GenBank+EMBL+ DDBJ EST Division	~~ftp:/~~/ncbi.nlm.nih.gov/ blast/est_human.Z	Expect value (e) < 10^{-25}
UniGene_unique	Build 132	One sequence selected from each UniGene cluster (the one with the longest region of high-quality sequence data).	~~ftp:/~~/ncbi.nlm.nih.gov/ pub/shuler/unigene/ Hs.seq.uniq.Z	Expect value (e) < 10^{-25}
Human Genome	Build 22	Sequence data of all contigs used to assemble the human genome	~~ftp:/~~/ncbi.nlm.nih.gov/ genomes/H_sapiens/ CHR_#/hs_chr#.fa.gz	Expect value (e) < 10^{-25}

On page 344, replace Table 6 with the following amended Table 6:

Table 6: Algorithms used for exon and polypeptide prediction

Algorithm	Description	Web address
Genscan	Predicts the locations and exon-intron structures of genes in genomic sequences.	~~http:/~~/genes.mit.edu/GENSCAN.html
Genomescan	Incorporates protein homology information when predicting genes.	~~http:/~~/genes.mit.edu/genomescan.html
GrailEXP	Predicts exons, genes, promoters, polyAs, CpG islands, EST similarities, and repetitive elements within a DNA sequence.	~~http:/~~/grail.lsd.ornl.gov/grailexp[[/]]
G-Known	Predicts genes and features of a DNA sequence at user-specified levels of complexity. Can incorporate extra information supplied by user including gene predictions from other gene finding programs, EST hits, similarities to known proteins, synteny between corresponding genomic regions in related organisms, methylation of the bases, regulatory binding sites, and topology information.	~~http:/~~/www.cse.ucsc.edu/research/compbio/pgf[[/]]
FGENES	Uses linear and hidden Markov models for exon prediction	~~http:/~~/genomic.sanger.ac.uk/gf/gf.shtml

On page 345, replace Table 7 with following amended Table 7:

Table 7: Databases and algorithms used for Protein Analysis

Algorithm	**Description**	**Web address**
BLASTP, version 2.0	Identification of unknown protein or subunit based on similarity to known proteins or subunits.	~~http:~~//www.ncbi.nlm.nih.gov/BLAST [[/]]
BLASTX	Algorithm for translating a nucleotide query sequence and aligning the translation to sequences in protein databases	~~http:~~//www.ncbi.nlm.nih.gov/BLAST [[/]]
TBLASTN	Algorithm for aligning an unidentified peptide sequence to predicted translations of nucleotide sequences	~~http:~~//www.ncbi.nlm.nih.gov/BLAST [[/]]
SWISS-PROT, release 39.0	Protein sequence database	~~http:~~//www.expasy.ch/cgi-bin/ sprot-search-de
Protein International Resource (PIR)	Protein sequence database	~~http:~~//www-nbrf.georgetown.edu/ pirwww[[/]]
GenPept	Amino acid translations from GenBank/EMBL/DDBJ records that are annotated with one or more CDS features	~~ftp:~~//ncbi.nlm.nih.gov/genbank/ genpept.fsa.gz
TrEMBL	Contains the translations of all coding sequences present in the EMBL Nucleotide Sequence Database, which are not yet integrated into SWISS-PROT	~~http:~~//www.ebi.ac.uk/swissprot[[/]]
Prosite, release 16.39	Database of protein families and domains. Consists of biologically significant sites, patterns and profiles.	~~http:~~//www.expasy.ch/prosite[[/]]
Pfam, version 6.2	Collection of multiple sequence alignments and hidden Markov models covering many common protein domains	~~http:~~//www.sanger.ac.uk/Software/ Pfam[[/]]

ProDom, version 2001.1	Domain arrangements of proteins and protein families	~~http:/~~/protein.toulouse.inra.fr/ prodom.html
TMpred	Prediction of transmembrane regions to aid in subcellular localization and function predictions	~~http:/~~/www.ch.embnet.org/software/ TMPRED_form.html